



De l'usage des logiques modales pour la gestion de l'incertitude des données : application en archéologie

Elisabeth Gavignet, Eric Leclercq, Nadine Cullot, Marinette Savonnet

► To cite this version:

Elisabeth Gavignet, Eric Leclercq, Nadine Cullot, Marinette Savonnet. De l'usage des logiques modales pour la gestion de l'incertitude des données : application en archéologie. SAGEO'2015 (Spatial Analysis and GEomatics), GdR MAGIS, Nov 2015, Hammamet, Tunisie. hal-01244611

HAL Id: hal-01244611

<https://hal.science/hal-01244611>

Submitted on 16 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

De l'usage des logiques modales pour la gestion de l'incertitude des données : application en archéologie

Élisabeth Gavignet, Éric Leclercq, Nadine Cullot, Marinette Savonnet

*Laboratoire LE2I - UMR6306 - CNRS - ENSAM
Univ. Bourgogne Franche-Comté
9, Avenue Alain Savary
F-21078 Dijon - France
Prenom.Nom@u-bourgogne.fr*

RÉSUMÉ. Les systèmes d'information archéologiques offrent des méthodes et des outils pour représenter les données collectées et effectuer des analyses avec lesquels la prise en compte de données imparfaites est souvent délicate. Notre contribution porte sur l'utilisation de plusieurs logiques modales pour modéliser et tester les effets de la prise en compte de données incertaines, mais aussi pour vérifier la qualité d'un corpus tout en travaillant dans un environnement collaboratif interdisciplinaire. La modélisation et le raisonnement sur des données incertaines, proposés dans cet article, s'intègrent dans une plateforme ouverte et extensible permettant de gérer les données archéologiques. Sur le plan informatique, le raisonneur utilisé, basé sur la logique du premier ordre, fournit aux archéologues, les moyens de valider ou non leurs hypothèses. Quelques exemples de raisonnement sont proposés pour des données décrivant un personnage ecclésiastique.

ABSTRACT. Archaeological information systems offer methods and tools for representing collected data and performing analyses with which taking into account imperfect data is often hard to please. Our contribution describes the use of several modal logics to model and verify the effects of the consideration of uncertain data, but also to check the quality of a corpus in an interdisciplinary collaborative environment. The modelling and the reasoning based on uncertain data, which are studied in this article, are integrated open and extensible platform allowing to manage archaeological data. From the computing point of view, the reasoner used, based on the first order logic, provides the archaeologists with ways to validate or not their hypotheses. Some examples of reasoning are proposed for data describing an ecclesiastical person.

MOTS-CLÉS : incertitudes, logiques modales, raisonnement, annotations sémantiques, ontologies
KEYWORDS: uncertainty, modal logics, reasoning, semantic annotations, ontology

1. Introduction

Les recherches en archéologie sont de plus en plus interdisciplinaires et collaboratives. Les données proviennent de sources multiples (fouilles, données cartographiques, sources historiques, etc.) et leur collecte a recours à différentes techniques (de datation, de caractérisation des matériaux, etc.) opérées par de nombreux spécialistes. Elles sont exploitées dans des cadres variés : par des spécialistes de la restauration des bâtiments ou par des organismes du ministère de la culture, pour des missions d'inventaire, d'étude, de protection ou de diffusion du patrimoine archéologique.

L'ensemble de ces acteurs ont, par conséquent, besoin d'une plateforme qui va bien au delà d'une simple base de données. Elle doit proposer des solutions pour répondre à deux problématiques : la première est l'extensibilité des modèles de données sans remettre en cause l'opérabilité des applicatifs et la seconde est la préservation de la qualité des données et des connaissances. Pour cela, l'archéologie a donc besoin de systèmes évolutifs, d'outils de détection d'incohérence (Leclercq, Savonnet, 2012) et de modèles de représentation et d'interprétation de données incertaines.

La multiplicité des acteurs impliqués accroît la difficulté d'obtenir un consensus sur les connaissances représentées. Les ontologies proposent une spécification formelle et explicite d'une conceptualisation partagée d'un domaine (Gruber, 1993). À ce titre, elles sont un moyen facilitant la représentation des données. Les raisonneurs qui leur sont associés peuvent prendre en charge la détection d'incohérences tout en inférant de nouvelles connaissances. La modélisation des ontologies, généralement basée sur des langages logiques comme OWL (Ontology Web Language), n'offrent pas d'opérateurs adaptés à la représentation des connaissances imparfaites. Seuls des systèmes plus complexes mixant plusieurs paradigmes pour la modélisation des connaissances permettent de représenter et d'exploiter cette variété d'information.

La suite de l'article est organisée comme suit. La section 2 présente un état de l'art sur les typologies de l'incertitude, montre l'intérêt des logiques modales pour modéliser certaines formes de connaissances incertaines et s'intéresse aux méthodes de raisonnement sur ces connaissances. La section 3 décrit l'architecture de la plateforme WikiBridge, puis développe l'utilisation des logiques modales pour trois cas distincts de raisonnement : la gestion de la localisation, la datation et l'interprétation multi-agents. La section 4 dresse le bilan de nos travaux et présente des perspectives plus larges d'usages des logiques modales comme base pour la modélisation et le raisonnement sur des données incertaines dans les systèmes d'information scientifiques.

2. État de l'art

Dans cette partie nous dressons un panorama des classifications de l'incertitude et étudions quels outils de modélisation et de raisonnement peuvent être utilisés pour traiter des données incertaines.

2.1. Typologies des incertitudes

Dans le dictionnaire Larousse l'incertitude est définie comme « l'état de quelqu'un qui ne sait pas quel parti prendre ». Plusieurs classifications de l'incertitude ont été proposées, il n'est pas dans notre intention d'en donner un inventaire complet mais d'en citer quelques unes qui nous semblent avoir des points de vue intéressants mais pas forcément communs.

Dans le domaine des systèmes d'information géographique, (Goodchild *et al.*, 1994) associent l'incertitude à quatre notions : l'exactitude, la validité, la qualité et le bruit. (Gershon, 1998) considère l'incertitude comme l'une des six causes de l'imperfection au même titre que l'incohérence, l'incomplétude, les données erronées, les données complexes et les présentations imparfaites. (Thomson *et al.*, 2005) proposent une répartition de neuf critères d'incertitude dans trois catégories : qualité de l'information, cohérence de l'information et objectivité de l'information (tableau 1). Suite à

Tableau 1. Synthèse de la typologie de l'incertitude de Thomson *et al.*, 2005

Qualité	Cohérence	Objectivité
exactitude de la collecte	concordance entre données	fiabilité des sources
précision de la mesure	transformation	subjectivité
exhaustivité des informations	lacune temporelle	indépendance de la source

une enquête pluridisciplinaire (Skeels *et al.*, 2010) décrivent une classification en trois niveaux hiérarchiques et deux niveaux transversaux. Le niveau le plus faible d'incertitude est relatif à la précision de la mesure, le niveau moyen fait référence à la complétude des informations (celles dont on dispose, celles dont on ne dispose pas mais dont on connaît l'existence et celles dont on ignore l'existence) et le niveau le plus élevé correspond aux données produites par des inférences. Quant aux niveaux transversaux, ils concernent les notions de désaccord et de crédibilité. (Fisher, 1999) propose, à partir du modèle conceptuel d'incertitudes de (Klir, Yuan, 1995), une version adaptée qui considère le flou et l'ambiguïté d'objets pas bien définis comme un secteur où aucune théorie de représentation des données n'avait été jugée satisfaisante. (Smets, 1997) a établi une classification de l'imperfection éclatée en trois principaux groupes : imprécision, incohérence et incertitude. Pour l'auteur, l'imprécision est relative au contenu des déclarations qui peuvent contenir ou non des erreurs. L'incohérence regroupe les déclarations conflictuelles. Quant à l'incertitude, elle est induite soit par des informations objectives (les propriétés de la donnée) soit par des informations subjectives (l'avis de l'observateur). La classification de (De Runz, 2008), dans le domaine archéologique, n'est pas si éloignée de celle proposée par Smets. Cependant, elle place l'ambiguïté au même niveau que l'incertitude alors que Smets la considère comme une forme d'imprécision. Cette distinction est tout à fait légitime dans la mesure où De Runz propose cette classification comme support au choix de la théorie de représentation la mieux adaptée pour modéliser l'information. Il conforte l'idée selon laquelle l'ambiguïté reste délicate à traiter. Elle relève à la fois de l'imprécision et de l'incertitude. Toujours dans le domaine de l'archéologie, une classification a été proposée à l'occasion de la première table ronde « L'incertitude : regards pluridisci-

plinaires » organisée par F. Bertoncello en 2013. Elle distingue les données lacunaires, les données ambiguës et les données imprécises. À ces catégories d'incertitude liées aux données viennent s'ajouter des incertitudes liées à l'analyse ou à l'interprétation qui est faite de ces dernières.

Chacune de ces classifications reste discutable et selon le domaine, certaines correspondent mieux à une réalité que d'autres. Néanmoins, on peut remarquer plusieurs similitudes. En particulier, la cohérence (respectivement l'incohérence) se retrouve dans une majorité de classifications avec parfois des variantes au niveau du vocabulaire. Il semblerait aussi que les données issues du raisonnement humain (l'interprétation, la prédiction, l'extrapolation, etc.) soient davantage incertaines que les données mesurées. À l'inverse, la qualité de l'information apparaît comme telle dans deux classifications alors qu'elle est déclinée dans les autres à la manière de ce qui est fait dans le tableau 1. Tous les travaux soulignent l'impact de l'incertitude sur la qualité des données et des décisions qui sont prises. Sa représentation ne doit pas être minorée au risque de remettre en question la validité des résultats inférés. L'utilisation d'un cadre théorique formel, comme les logiques modales, et la spécification des mécanismes de raisonnement associés constituent une problématique importante.

2.2. De la logique classique aux logiques modales

Dans ce paragraphe, nous définissons les fondamentaux de la logique des prédicats du premier ordre puis nous détaillons les concepts des logiques modales utiles à la compréhension de cet article.

La **logique des prédicats** du premier ordre permet de représenter des connaissances à l'aide de formules basées sur la définition de prédicats n -aires, l'utilisation d'opérateurs logiques (conjonction \wedge , disjonction \vee , négation \neg , implication \rightarrow) et de quantificateurs. Une formule composée uniquement d'un prédicat est une formule atomique ou atome. Les formules non atomiques sont construites à partir d'autres formules, des opérateurs logiques et des quantificateurs universel (\forall) et existentiel (\exists). Le calcul des prédicats s'intéresse d'une façon générale à déterminer la véracité d'une formule. Il n'est généralement pas décidable, aussi est-il souhaitable de restreindre la description des formules à des formes permettant la mise en œuvre de moteurs de raisonnement complet, capables de prouver la véracité ou non d'une formule. Cette restriction consiste à travailler sur une forme clausale particulière : les clauses de Horn. Une clause de Horn a la forme suivante : $B_1, \dots, B_n \rightarrow H$ où B_1, \dots, B_n représente le corps de la clause et H la tête. Chacun des symboles H, B_i désigne un prédicat. Une clause dont le corps est vide énonce un fait. Une clause avec une tête et un corps non vides traduit une règle. Cette dernière indique que si les conditions B_1 et ... et B_n sont vraies alors la conclusion H est aussi vraie. La forme logique d'une telle clause est $\forall x_1, \dots, x_k (H \vee \neg B_1 \vee \dots \vee \neg B_n)$ où x_i désigne les variables référencées dans les prédicats.

Les **logiques modales** permettent, par l'introduction de nouveaux opérateurs, appelés modaux (le nécessaire \Box et son dual le possible \Diamond), d'enrichir leur pouvoir d'ex-

pression et de modéliser des formules plus nuancées. On peut par exemple, vouloir décrire la proposition « Il est possible que l'église Notre Dame de Dijon ait été une simple chapelle » à l'aide d'une modalité de *possibilité*, ou encore, utiliser la modalité *on sait que* pour affirmer « On sait que l'église Notre Dame de Dijon a été une simple chapelle ».

Il existe plusieurs catégories de logiques modales classées selon les interprétations des opérateurs modaux définis dont les principales sont : logique aléthique (possible ou nécessaire), temporelle (parfois ou toujours), épistémique (croit que, sait que), déontique (est permis, est obligatoire), etc. La compréhension de ces modalités nécessite une extension de la sémantique de la logique des prédicats du premier ordre qui est classiquement basée sur des modèles de Kripke (Gasquet *et al.*, 2013). Ces derniers considèrent l'existence de différents mondes reliés entre eux par une relation d'accessibilité. Ainsi, en logique aléthique, une formule φ pourra être interprétée comme *possible*, notée $\Diamond\varphi$, si elle est vraie dans un monde et fausse dans un autre et comme *nécessaire*, notée $\Box\varphi$, si elle est vraie dans tous les mondes.

Les **systèmes de déduction** s'appuient sur les axiomes de la logique considérée pour réaliser des déductions à partir des formules décrites et connaître la véracité d'une formule. Ils utilisent les mécanismes de substitution pour remplacer une variable par une autre expression et d'unification qui est une succession de substitutions rendant deux formules égales.

La méthode de résolution SLD (Selective Linear Definite) (Lloyd, 1987) est une méthode de preuve complète de la logique des prédicats du premier ordre sur un ensemble P de clauses de Horn. La preuve est réalisée par réfutation. Pour prouver que le but F est vrai, il faut prouver que $P \cup \{\neg F\}$ est insatisfiable. Soit $G = \neg F$, la démonstration est faite en dérivant à partir de G une séquence de formules jusqu'à obtenir la clause vide. Les formules G_i sont telles que $G_0 = G$ et pour tout $i \geq 0$, G_{i+1} est obtenu à partir de G_i , de la manière suivante :

- A_m est un atome sélectionné dans le corps de G_i
- $B_1, \dots, B_k \rightarrow H$ est une clause du programme P
- θ est l'unificateur le plus général de A_m et H
- $G_{i+1} = \rightarrow (A_1, \dots, A_{m-1}, B_1, \dots, B_k, A_{m+1}, \dots, A_n)\theta$

Ce principe peut être étendu pour raisonner sur des formules de la logique modale. Plusieurs approches ont été proposées pour ce faire, comme la méthode des tableaux sémantiques (Fitting, 1988), la méthode des séquents de Gentzen (Szabo, 1969) ou la programmation logique. La méthode des tableaux sémantiques est comme la SLD résolution, une méthode de preuve par réfutation. La négation de la formule à prouver est décomposée en sous-formules de sorte que si la formule est satisfaite alors les sous-formules qui la composent le sont aussi. La décomposition permet la construction d'un arbre de résolution et s'effectue par application de règles d'expansion (de prolongation, de ramification, de double négation, etc.) qui tiennent compte de la sémantique des mondes possibles. Une formule a une preuve si toutes les branches de l'arbre sont fermées, c'est-à-dire qu'une formule et sa négation apparaissent dans chaque branche.

En programmation logique, les approches consistent à étendre les langages de programmation logique comme Prolog, pour prendre en compte les opérateurs modaux. Elles sont détaillées dans la section 3 et notamment le langage MProlog (Nguyen, 2004) utilisé dans notre travail.

3. Approche proposée

Les données archéologiques sont complexes à gérer car elles représentent des volumes de données hétérogènes importants. De plus elles possèdent un grand nombre d'attributs, évoluent avec le temps et proviennent de plusieurs sources. La constitution et la publication d'un corpus dans un environnement collaboratif est une demande forte dans le domaine de l'archéologie. Le nombre important des acteurs impliqués pose néanmoins le problème de la qualité des informations collectées et leur concordance les unes avec les autres. Ce sont ces raisons qui ont conduit au développement d'une plateforme WikiBridge (Chevalier *et al.*, 2010) et à la mise en place de moyens pour contrôler la qualité des informations collectées.

3.1. Présentation de l'architecture

Le cœur de WikiBridge est constitué par le couplage entre des bases de données relationnelles et des annotations sémantiques pour produire des connaissances dans un environnement collaboratif évolutif. Ce couplage permet de réaliser une extensibilité dynamique des schémas de données tout en conservant les mécanismes de contrôle et le fonctionnement des applicatifs. L'interface utilisateur prend la forme d'un wiki pour décrire des objets, des édifices, des lieux au moyen d'une description riche incluant textes, sources bibliographiques, plans, localisations, photographies. Les différentes ressources qui constituent les descriptions des éléments archéologiques peuvent être complétées par des annotations construites à partir d'une ontologie sous la forme de triplets < sujet, prédicat, objet > (Leclercq *et al.*, 2013). Ces annotations sont soumises à deux types de vérification de la consistance : l'une logique vérifie que les propriétés associées aux types d'objets annotés sont cohérentes avec la description ontologique, l'autre sémantique vérifie la cohérence des annotations entre elles (non contradiction) et la cohérence des annotations par rapport aux règles du domaine (connaissance spécifique non nécessairement consensuelle).

La figure 1 présente les principaux modules de WikiBridge et leurs interactions. La saisie ainsi que la mise en forme sont implantées en utilisant une extension de MediaWiki avec les composants sémantiques suivants : une interface d'acquisition de données basée sur des formulaires permettant l'ajout de données structurées au schéma bien établi (dans une base de données relationnelle) et leur annotation sémantique automatisée à l'aide d'un assistant, un moteur de requêtes SPARQL, un mécanisme de contrôle des annotations incluant la prise en compte du contexte et des contraintes sémantiques. Pour l'analyse spatiale, un web service offre une interaction avec des logiciels externes et un module de cartographie s'appuyant sur OpenLayers. L'inter-

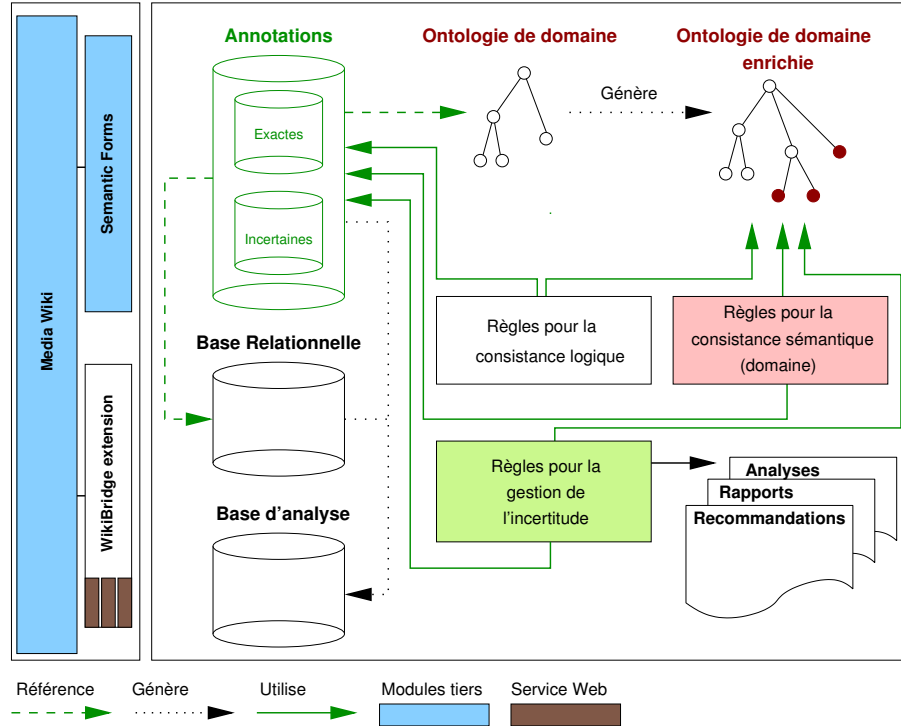


Figure 1. Architecture de WikiBridge

façage avec un SIG est réalisée grâce à une base de données spécifique (dite base d'analyse) dont le schéma est établi à partir du schéma relationnel et des annotations.

L'objectif des travaux que nous décrivons est de compléter les outils de raisonnement fournis par la plateforme¹ avec des outils de raisonnement sur des données et connaissances incertaines. En effet, l'incertitude peut se manifester à différents niveaux dans la production de connaissance et en particulier apparaître à la suite d'une détection d'incohérence.

3.2. Raisonnement en programmation logique modale

De nombreux travaux ont proposé des extensions de la programmation logique avec des approches différentes mais seuls le groupe de Fariñas del Cerro avec le système Molog (Del Cerro, 1986) et Linh Anh NGuyen avec MProlog (Nguyen, 2003) ont abouti à des implémentations. Si les deux systèmes optent pour la même démarche,

1. développée puis validée dans le cadre de deux projets : l'ANR CARE *Corpus Architecturae Religiosae Europaeae - IV-X saec.* - ANR-07-CORP-011 (<http://care.tge-adonis.fr>) et le projet de Plateforme numérique et mutualiste sur l'usage de la pierre bourguignonne (<http://lapierre.u-bourgogne.fr>).

celui de Nguyen n'impose aucune restriction quant à l'utilisation des opérateurs modaux. Il est basé sur le modèle de Kripke et intègre le calcul de résolution SLD. Ce modèle étendu prend en compte l'axiome de base (K) ainsi que d'autres axiomes (tableau 2) enrichissant le système de base (K-système). Les nouveaux systèmes de logique, obtenus par combinaison des axiomes avec l'axiome (K), portent généralement comme nom les initiales des axiomes (par exemple KD pour K + D).

Tableau 2. Axiomes caractérisant des relations entre les mondes de Kripke

Axiome	Formule	Propriété	Relation R entre mondes de Kripke ω_i
D	$\Box\varphi \rightarrow \Diamond\varphi$	sérielle	$\forall\omega_0 \exists\omega_1 R(\omega_0, \omega_1)$
T	$\Box\varphi \rightarrow \varphi$	réflexive	$\forall\omega_0 R(\omega_0, \omega_0)$
B	$\varphi \rightarrow \Box\Diamond\varphi$	symétrique	$\forall\omega_0, \omega_1 R(\omega_0, \omega_1) \wedge R(\omega_1, \omega_0)$
4	$\Box\varphi \rightarrow \Box\Box\varphi$	transitive	$\forall\omega_0, \omega_1, \omega_2 (R(\omega_0, \omega_1) \wedge R(\omega_1, \omega_2)) \rightarrow R(\omega_0, \omega_2)$
5	$\Diamond\varphi \rightarrow \Box\Diamond\varphi$	euclidienne	$\forall\omega_0, \omega_1, \omega_2 (R(\omega_0, \omega_1) \wedge R(\omega_0, \omega_2)) \rightarrow R(\omega_1, \omega_2)$

Un programme MProlog est un ensemble de formules restreintes aux clauses de Horn. Une clause avec modalité est une formule de la forme $[\Delta] : \varphi$ où Δ est une liste d'opérateurs modaux séparés par des virgules et φ une clause de Horn pouvant comporter des opérateurs modaux.

Les formules proposées dans le tableau 3 illustrent des possibilités d'expression des modalités et montrent qu'un opérateur modal peut indifféremment porter sur une clause ou sur un prédicat, dans le corps ou la tête de la clause.

Tableau 3. Exemples de modalités avec MProlog

ref	formule avec modalités	traduction en MProlog
φ_1	$\Diamond m(jane, alain)$	$[d]:m(jane, alain).$
φ_2	$\Box(\Box f(x) \leftarrow m(x, y))$	$[b]:([b]:f(X):- m(X, Y)).$
φ_3	$\Box(p(x, y) \leftarrow m(x, y), \Box f(x))$	$[b]:(p(X, Y) :- m(X, Y), [b]: f(X)).$
φ_4	$\Box e(x) \leftarrow \Diamond p(y, x)$	$[b]:e(X) :- [d]:p(Y, X).$

Prenons un exemple pour illustrer les étapes du processus de résolution du but $G = \leftarrow \Box e(x)$ à partir du programme P constitué des formules du tableau 3 dans la logique KD. En considérant G et φ_4 nous dérivons un nouveau but $G_1 = \leftarrow \Diamond p(y, x)$. À partir de G_1 et φ_3 , nous déduisons le but $G_2 = \leftarrow \Diamond (m(y, x) \wedge \Box f(y))$. Or G_2 n'est pas correct car la modalité porte sur une formule non atomique. Pour résoudre ce problème, le but $\Diamond p(y, x)$ est d'abord étiqueté par $\langle X \rangle p(y, x)$ où X est une variable libre (non quantifiée). Ainsi les formules φ_3 et K peuvent être appliquées à ce but étiqueté pour donner le nouveau but $G'_2 = \leftarrow \langle X \rangle m(y, x) \wedge \langle X \rangle \Box f(y)$. Nous pouvons résoudre G'_2 avec φ_1 puisque cette formule peut être traitée comme $\langle m(jane, alain) \rangle m(jane, alain)$. Cette résolution de G'_2 avec φ_1 fournit $G_3 = \langle m(jane, alain) \rangle \Box f(jane)$ grâce à l'unificateur $x/alain, y/jane, X/m(jane, alain)$. De G_3 et φ_2 nous obtenons $G_4 = \leftarrow \langle m(jane, alain) \rangle m(jane, y)$. En appliquant φ_1 à G_4 nous dérivons $G_5 = \leftarrow \langle m(jane, alain) \rangle m(jane, alain)$. Et finalement, en unifiant G_5 et φ_1 nous arrivons à la clause vide.

Cette démonstration ne relate pas en détail tous les mécanismes mis en œuvre par

MProlog pour résoudre le but mais l'essentiel est fourni pour comprendre les cas d'utilisation qui seront traités.

3.3. Cas d'utilisation de différentes formes d'incertitude

La complexité intrinsèque de l'incertitude rend très difficile la généralisation d'une modélisation universelle. Plusieurs travaux ont utilisé ces logiques modales pour traiter de données aussi variées que la sécurité des informations personnelles (Piolle, Demazeau, 2008) en logique déontique, les incidences géométriques (Balbiani *et al.*, 1997) avec des relations temporelles, la réglementation dans des systèmes multi-agents (Garion *et al.*, 2010) reformulée dans un cadre modal du premier ordre, la définition de régions dans l'espace et de raisonnement spatial (Bentham, Bezhanishvili, 2007 ; Bennett, 1996) mais aucun, à notre connaissance, n'a fait l'objet d'une implémentation en programmation logique modale.

Notre contribution porte sur la possibilité de disposer d'un formalisme basé sur la logique modale et la logique des prédicats qui modélise certaines formes d'incertitude afin de rendre compte des effets de ces données incertaines sur l'ensemble des données existantes. Nous considérons l'incertitude liée d'une part aux propriétés des données représentées et d'autre part à l'interprétation que l'observateur peut en faire.

L'exemple choisi (figure 2) est tiré de fiches ressources qui ont été proposées lors des ateliers de l'école thématique du GdR Modys². Il regroupe un ensemble d'imperfections propices à la représentation des modalités souhaitées. Dans cette fiche, les informations considérées comme certaines sont : (a) St Léontius est évêque de Fréjus en Provence, (b) il est né vers la fin du 4^e siècle, (c) il est mort dans sa ville épiscopale, (d) sa fête est le 1^{er} décembre. Il est important d'attirer l'attention sur le fait que l'extrait ne précise pas qu'il est mort à Fréjus. Cependant cette information pourra être déduite du fait que Léonce est évêque de Fréjus et que la ville où un évêque exerce sa fonction correspond à sa ville épiscopale. Concernant les informations incertaines, seule celle relative au lieu de naissance est concernée : (e) il est probablement né à Nîmes. Enfin, la présence de trois dates différentes pour la mort de Léonce de Fréjus montrent des contradictions. Plusieurs options sont envisageables : ne modéliser aucune donnée, modéliser une seule d'entre elles ou mémoriser toutes ces données. C'est cette dernière option que nous retiendrons. Le stockage de plusieurs valeurs pour une caractéristique usuellement monovaluée pose problème dans bon nombre de représentations. Dans une base de données relationnelle cette difficulté peut être contournée mais systématiser la solution à tous les attributs ne serait pas pertinent. Au niveau de l'ontologie, si la propriété a été définie comme fonctionnelle une incohérence sera automatiquement détectée. En logique des prédicats du premier ordre il est possible d'énoncer trois dates mais disposer de trois valeurs pour un même événement relève d'un cas d'incertitude. Il y a forcément des valeurs non correctes parmi celles

2. Modélisation et visualisation des dynamiques spatiales : Reasonner sur le temps long et ses incertitudes
http://www.map.archi.fr/modys/resultats_fr.htm

fournies et tout au plus une valeur correcte. C'est à partir de ces informations que nous proposons d'étudier deux modalités (aléthique et épistémique) pour illustrer une modélisation différenciée d'une même source.

Leontius of Fréjus

From Wikipedia, the free encyclopedia

Saint Leontius (**French**: *Léonce de Fréjus*) (ca. 419 – ca. 488) was a **bishop of Fréjus**, in **Provence**. He was born probably at **Nîmes**, towards the end of the fourth century; he died in his episcopal town in 488, according to some authorities, though others say 443 or even 448. His **feast day** is December 1.

Figure 2. Exemple extrait d'une fiche ressource

Les annotations qui complètent les descriptions des éléments archéologiques sont transcrites sous forme de triplets. Par exemple, le triplet ('Fréjus', estLocaliseEn, 'Provence') exprime le fait que "Fréjus est en Provence". Lors du traitement par le système MProlog, une forme clausale vient se substituer au triplet. Celle-ci est constituée du prédicat d'arité trois *an* (abrégé de annotation) et des valeurs présentes dans le triplet. Ce qui donne *an('Fréjus', estLocaliseEn, 'Provence')* avec le triplet précédemment utilisé comme exemple.

La première solution est l'utilisation de la **logique modale aléthique**. Dans ce cas l'ensemble des annotations sont considérées comme fournies par la même personne à partir de plusieurs sources ou par plusieurs personnes à partir de plusieurs sources sans qu'il soit utile de différencier la provenance. Les modalités qualifient les données par rapport à leur véracité sans remise en cause de la personne qui fournit l'information. Le tableau 4 propose la traduction des propositions, concernant Léonce de Fréjus, classées dans le même ordre que sur la fiche (Figure 2). Afin de réduire la longueur des formules, stLeontius est abrégé en stL.

Tableau 4. Formules en logique modale aléthique

	Formule	Proposition
(f1)	$\text{an}(\text{stL}, \text{assureFct}, \text{eveque})$	(a)
(f2)	$\text{an}(\text{stL}, \text{aPourVilleEpiscopale}, 'Fréjus')$	
(f3)	$\text{an}('Fréjus', \text{estLocaliseEn}, 'Provence')$	
(f4)	$\Diamond \text{an}(\text{stL}, \text{estNeA}, 'Nîmes')$	(e)
(f5)	$\Box \text{an}(\text{stL}, \text{estNePdt}, \text{fin4s})$	(b)
(r1)	$\text{an}(\text{stL}, \text{aPourVilleEpiscopale}, V) \rightarrow \text{an}(\text{stL}, \text{estMortA}, V^a)$	(c)
(f6)	$\Diamond \text{an}(\text{stL}, \text{estMortEn}, 488)$	contradictions
(f7)	$\Diamond \text{an}(\text{stL}, \text{estMortEn}, 443)$	
(f8)	$\Diamond \text{an}(\text{stL}, \text{estMortEn}, 448)$	
(f9)	$\text{an}(\text{stL}, \text{aFete}, '01-12')$	(d)

a. V représente une variable : la ville où est mort stL et qui est aussi sa ville épiscopale.

Pour disposer d'un ensemble de formules plus riche permettant davantage de résolutions, nous proposons de rajouter deux règles :

- (r2) : $\text{an}(\text{Suj}, \text{Pred}, \text{Obj}) \rightarrow \Box \text{an}(\text{Suj}, \text{Pred}, \text{Obj})$
- (r3) : $\Box (\text{an}(\text{Suj}, \text{estMortEn}, D_m), \text{an}(\text{Suj}, \text{estNeEn}, D_n),$
 $\text{Age} = D_m - D_n \rightarrow \text{an}(\text{Suj}, \text{estMortA}, \text{Age}))$

La règle r2 permet de traiter les formules sans modalité comme des formules toujours nécessairement vraies. La règle se lit « Si la formule est vraie alors il est nécessaire qu'elle soit vraie ». Ce qui est tout à fait pertinent par rapport à l'interprétation qui a été proposée car un seul point de vue est considéré.

La règle r3 déduit l'âge d'une personne (Suj) au moment de son décès en fonction de son année de mort (D_m) et de son année de naissance (D_n) quand ces dernières sont connues. Il convient de remarquer que la modalité *nécessaire* (\Box) s'applique à l'ensemble de la formule et non uniquement au terme qui suit comme cela est le cas dans r2. Cette formulation présente l'avantage de permettre la propagation de l'incertitude. Lorsque l'une des formules de la partie gauche sera incertaine, la formule déduite le sera aussi. C'est d'ailleurs ce qui se passera avec les trois valeurs de la date de décès de Léonce de Fréjus qui vont générer trois nouvelles formules incertaines dès lors que le fait que « Léonce de Fréjus est né en 419 » est connu. La déduction selon laquelle il est mort à Fréjus sera une formule nécessairement vraie puisque déduite d'autres (f2, r1, r2) toutes nécessairement vraies. Avec cette modalité, toutes les données et les règles de déduction sont traitées comme provenant d'une seule et même personne (source). L'exemple des trois dates possibles issues de trois informations différentes montre à quel point la multiplication des doutes peut générer des déductions tout aussi incertaines. Dans ce cas, l'archéologue alerté de l'existence de données contradictoires pourra s'interroger sur leur véracité. La modélisation du monde réel conduit, certaines fois, à des cas où ce choix peut être délicat. Dans des situations particulières, quand les circonstances justifient un intérêt à conserver toutes les informations pour opérer des déductions différenciées, il est nécessaire de passer à une autre forme de logique : la logique multimodale $K_{(m)}$ avec m paires d'opérateurs modaux \Box_i et \Diamond_i .

La deuxième solution propose d'exploiter la **logique multimodale épistémique** puisque cette dernière est généralement utilisée dans le cas de système multi-agents. Les axiomes relatifs à ces logiques sont représentés dans le tableau 5. D'une certaine manière, c'est bien le point de vue des agents (sources) que nous cherchons à représenter et non plus les connaissances de manière globale indépendamment de leur source. La communauté des archéologues à laquelle nous nous intéressons collabore à la constitution de bases de connaissances communes au sein d'un groupe de confiance. Chacun veut pouvoir exprimer ses croyances et raisonner en exploitant celles des autres. Ayant besoin de représenter une croyance individuelle, nous avons retenu la logique $KD4_s5_s$ pour laquelle $\Box_i \varphi$ signifie « l'agent i croit que φ est vraie » et $\Diamond_i \varphi$ signifie « φ est considérée comme possible par l'agent i » où φ est une formule. Pour illustrer ce qui peut être fait avec cette logique, nous proposons d'exprimer le recoupement de trois sources différentes (S1, S2, S3) alimentant la connaissance sur le même sujet Léonce de Fréjus. La source S3 correspond au contenu de la fiche (figure 2). Nous avons attribué à S1 et S2 des croyances de manière à couvrir un maximum de combinaisons possibles. Par ailleurs, comme nous nous focalisons sur la repré-

Tableau 5. Extrait des axiomes de la logique multimodale

Nom	Règle	Commentaire
(D)	$\Box_i \varphi \rightarrow \neg \Box_i \neg \varphi$	consistance
(I)	$\Box_i \varphi \rightarrow \Box_j \varphi$ if $i > j$	i, j : degré de croyance
(4s)	$\Box_i \varphi \rightarrow \Box_j \Box_i \varphi$	introspection positive
(5s)	$\neg \Box_i \varphi \rightarrow \Box_j \neg \Box_i \varphi$	introspection négative

sensation des données incertaines, seul un sous-ensemble (f2, r1) des formules sans modalités du tableau 4 ainsi que deux des trois formules portant sur les dates (f6, f7), sont reprises ici.

Tableau 6. Formulation des données en $KD4_s5_s$

Ref	Formule	S1	S2	S3
(m1)	$\text{an}(\text{stL}, \text{estMortEn}, 488)$	\Box_1		\Diamond_3
(m2)	$\text{an}(\text{stL}, \text{estMortEn}, 443)$		\Box_2	\Diamond_3
(m3)	$\text{an}(\text{stL}, \text{estNeEn}, 419)$	\Box_1	\Box_2	\Box_3
(m4)	$\text{an}(\text{stL}, \text{estNeA}, \text{'Nimes'})$	\Box_1	\Box_2	\Diamond_3
(m5)	$\text{an}(\text{stL}, \text{aPourVilleEpiscopale}, \text{'Fréjus'})$	\Box_1	\Diamond_2	\Box_3
(r3)	$\text{an}(\text{S}, \text{estMortEn}, D_m), \text{an}(\text{S}, \text{estNeEn}, D_n),$ $\text{Age} = D_m - D_n \rightarrow \text{an}(\text{S}, \text{estMortA}, \text{Age})$	\Box_1	\Box_2	\Box_3
(m6)	$\text{an}(\text{stL}, \text{aPourVilleEpiscopale}, \text{V}) \rightarrow \text{an}(\text{stL}, \text{estMortA}, \text{V})$	\Diamond_1	\Diamond_2	\Diamond_3

La représentation des données extraites de la fiche donnerait l'ensemble des formules du tableau 6. Les deux premières lignes du tableau (m1, m2) décrivent que S1 et S2 savent chacun la date de mort de Léonce de Fréjus et que S3 ne peut qu'y croire. En effet, en présence des deux dates S3 ne sait pas laquelle des deux dates est la bonne. Pour lui les deux dates sont envisageables puisque S1 et S2 l'affirment. À l'inverse, toutes les sources sont d'accord pour confirmer que Léonce est né en 419 (m3) et que le calcul de l'âge du décès est la différence entre les années de mort et de naissance (r2). Quant à la règle relative au lieu de sa mort (m6), les trois sources émettent un doute.

Ainsi représentées, les données peuvent être consultées par rapport à leur crédibilité. À tout moment S3 peut consulter les données qu'il confirme, celles qui deviennent probables et les autres qui sont uniquement possibles. À titre d'illustration, la règle suivante : $\Box_3 (\text{an}(\text{Suj}, \text{Pred}, \text{Obj}), \text{for_all}(\text{I}, [1,2], \Box_I \text{an}(\text{Suj}, \text{Pred}, \text{Obj}))) \rightarrow \text{confirm}(\text{Suj}, \text{Pred}, \text{Obj})$ retrouve toutes les annotations confirmées par S3. Elle indique que si une annotation est considérée comme vraie pour S3 et aussi considérée comme vraie pour les sources S1 et S2 alors c'est une annotation confirmée. Seule l'annotation (m3) répond à ce critère. Une formule logique similaire pourrait être écrite pour retrouver les annotations probables suite à la prise en compte de S3 c'est-à-dire possibles pour S3 mais nécessairement vraies pour toutes les autres. Ce qui donnerait ici les annotations (m3) et (m4). Les annotations (m1) et (m2) ne font pas partie du résultat car elles sont effectivement possibles pour S3 mais ne sont nécessaires que dans une seule des autres sources. En logique modale, les annotations confirmées sont forcément des

annotations probables. En effet, ce qui est vrai partout est forcément vrai dans un des mondes. De façon analogue, une règle peut détecter toutes les annotations considérées comme possibles pour au moins une source.

4. Conclusion

La modélisation et le raisonnement sur des données incertaines sont des tâches complexes qui nécessitent l'usage de paradigmes multiples et d'outils particuliers permettant d'enrichir le pouvoir d'expression des langages ou modèles utilisés pour des données classiques. L'approche proposée dans cet article consiste à s'appuyer sur des logiques modales pour représenter certaines formes d'incertitudes. Nous avons retenu deux logiques particulières : la logique aléthique qui introduit les opérateurs modaux de *possibilité* et de *nécessité*, et la logique épistémique qui élargit la logique aléthique avec la notion d'agents qui peuvent avoir des croyances différentes. Le travail présenté s'intègre dans une plateforme collaborative de gestion de données archéologiques qui permet : 1) de décrire des données structurées dans une base de données relationnelle, des annotations sémantiques sur ces données contrôlées à l'aide d'une ontologie et la spécification de règles métiers pour la cohérence des données et 2) de raisonner sur ces données en s'appuyant sur des formalismes de la logique des prédicats du premier ordre pour les règles métiers du domaine et les logiques de description pour les annotations. Dans cet environnement multi-paradigme, les logiques modales sont utilisées pour représenter les connaissances incertaines mais aussi définir des règles de gestion de ces données incertaines permettant de faire émerger des connaissances utiles aux archéologues.

Le travail a été expérimenté sur des données dont l'incertitude était liée soit aux propriétés des données (localisation et datation) soit à l'interprétation des données par différents agents (Gavignet *et al.*, 2015). Il a permis d'envisager plusieurs perspectives de développement en relation avec différents points de la proposition. Le premier point est lié à la recherche de la bonne adéquation entre la logique modale utilisée et la forme des connaissances incertaines à modéliser pour répondre aux besoins. À court terme, nous proposons de tester les mécanismes de raisonnement en se basant sur les différentes modalités et les raisonnements multi-modaux offerts par MProlog. Le second point concerne le passage à l'échelle de la méthode proposée pour travailler sur des corpus complets et s'intéresser aux connaissances émergentes pour les archéologues. Les corpus en archéologie du bâti qui ont été précédemment modélisés avec les logiques classiques vont servir de base à ce travail.

Remerciements

Ces travaux ont été partiellement financés par le CNRS au travers du projet PEPS LogModArchéo en 2013 et 2014.

Bibliographie

- Balbani P., Cerro L. F. del, Tinchev T., Vakarelov D. (1997). Modal logics for incidence geometries. *Journal of Logic and Computation*, vol. 7, n° 1, p. 59–78.
- Bennett B. (1996). Modal logics for qualitative spatial reasoning. *Logic Journal of IGPL*, vol. 4, n° 1, p. 23–45.
- Bentham J. van, Bezhanishvili G. (2007). Modal logics of space. In *Handbook of spatial logics*, p. 217–298. Springer.
- Chevalier P., Leclercq E., Millereux A., Sapin C., Savonnet M. (2010). WikiBridge: a Semantic Wiki for Archaeological Applications. In *Proceedings of the 38th conference on computer applications and quantitative methods in archaeology (caa)*, p. 193–196.
- Del Cerro L. F. (1986). Molog: A system that extends prolog with modal logic. *New Generation Computing*, vol. 4, n° 1, p. 35–50.
- De Runz C. (2008). *Imperfection, temps et espace: modélisation, analyse et visualisation dans un sig archéologique*. Thèse de doctorat. Université de Reims Champagne-Ardenne.
- Fisher P. F. (1999). Models of uncertainty in spatial data. *Geographical information systems*, vol. 1, p. 191–205.
- Fitting M. (1988). First-order modal tableaux. *Journal of Automated Reasoning*, vol. 4, n° 2, p. 191–213.
- Garion C., Roussel S., Cholvy L. (2010). Une logique modale pour raisonner sur la cohérence et la complétude de réglementations. *Revue d'intelligence artificielle*, vol. 24, n° 3, p. 267–290.
- Gasquet O., Herzig A., Said B., Schwarzentruher F. (2013). *Kripke's worlds: An introduction to modal logics via tableaux*. Springer Science & Business Media.
- Gavignet E., Leclercq E., Cullot N., Savonnet M., Saligny L. (2015). *Présentation poster: First order logic and modalities for managing uncertainty in archaeological databases*.
- Gershon N. (1998). Visualization of an imperfect world. *Computer Graphics and Applications, IEEE*, vol. 18, n° 4, p. 43–45.
- Goodchild M., Battenfield B., Wood J. (1994). On introduction to visualizing data validity. *Visualization in geographical information systems*, p. 141–149.
- Gruber T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, vol. 5, n° 2, p. 199–220.
- Klir G., Yuan B. (1995). *Fuzzy sets and fuzzy logic* (vol. 4). Prentice Hall.
- Leclercq E., Savonnet M. (2012). Système d'information pour la production de connaissances. l'approche wiki sémantique. *Ingénierie des Systèmes d'Information*, vol. 17, n° 3, p. 143–166.
- Leclercq E., Savonnet M., Troya-Galvis A., Buttner S. (2013). Investigating a multi-paradigm system for the management of archaeological data: Corpus lapidum burgundiae. In *Digital heritage international congress (digitalheritage), 2013*, vol. 1, p. 679–682.
- Lloyd J. W. (1987). *Foundations of logic programming. symbolic computation: Artificial intelligence*. Springer-Verlag, 2nd, extended edition edition.

- Nguyen L. A. (2003). A fixpoint semantics and an sld-resolution calculus for modal logic programs. *Fundamenta Informaticae*, vol. 55, n° 1, p. 63–100.
- Nguyen L. A. (2004). The modal logic programming system mprolog. In *Logics in artificial intelligence*, p. 266–278. Springer.
- Piolle G., Demazeau Y. (2008). Une logique pour raisonner sur la protection des données personnelles. In *16e congrès francophone afrif-afia sur la reconnaissance de formes et l'intelligence artificielle (rfia'08)*.
- Skeels M., Lee B., Smith G., Robertson G. G. (2010). Revealing uncertainty for information visualization. *Information Visualization*, vol. 9, n° 1, p. 70–81.
- Smets P. (1997). Imperfect information: Imprecision and uncertainty. In *Uncertainty management in information systems*, p. 225–254. Springer.
- Szabo M. E. (1969). *The collected papers of gerhard gentzen* (vol. 160). North-Holland Amsterdam.
- Thomson J., Hetzler E., MacEachren A., Gahegan M., Pavel M. (2005). A typology for visualizing uncertainty. In *Electronic imaging 2005*, p. 146–157.